

Confidence bands in nonparametric errors-in-variables regression

Aurore Delaigle, Peter Hall and Farshid Jamshidi

Department of Mathematics and Statistics, University of Melbourne, VIC, 3010, Australia

Abstract: Errors in variables regression is important in many areas of science and social science, for example in economics where it is often a feature of hedonic models, in environmental science where air quality indices are measured with error, in biology where the vegetative mass of plants is frequently obscured by mismeasurement, and in nutrition where reported fat intake is typically subject to substantial error. To date, in nonparametric contexts, the great majority of work has focused on methods for estimating the mean as a function, with relatively little attention being paid to techniques for empirical assessment of the accuracy of the estimator. In this paper we develop methodologies for constructing confidence bands. Our contributions include techniques for tuning parameter choice aimed at minimising the coverage error of confidence bands.

Key words: Bandwidth choice, bootstrap, confidence interval, coverage accuracy, double bootstrap, kernel methods, measurement error, nonparametric curve estimation, nonparametric regression, pointwise confidence band, SIMEX methods.

Short title: Confidence bands

1 Introduction

Confidence bands are a particularly informative way of quantifying the accuracy of estimators of regression means. For example, they provide a graphically simple description of the way in which estimation accuracy varies with location. A lot of work has been devoted to the construction of such bands in the standard nonparametric regression problem, where the covariates are observed accurately. In this paper, we construct confidence bands in the nonparametric errors-in-variables regression problem, where explanatory variables are observed with error.

This errors-in-variables problem has received a lot of attention in the literature. In a nonparametric context, most of the work so far has concentrated on the development of consistent regression estimators. Some authors have developed confidence bands in related,

but simpler, deconvolution problems; see for example Bissantz et al. (2007), Bissantz and Birke (2009) and Birke et al. (2010). However, despite their practical importance, to our knowledge confidence bands in errors-in-variables nonparametric regression have largely been ignored so far. We show that the problem is particularly complex, much more so than in the standard error-free setting. We argue that methods that require explicit variance estimation, such as confidence bands based on central limit theory or the percentile- t bootstrap, are particularly difficult to implement in practice, and we focus instead on percentile methods, which estimate variance implicitly.

We develop our approach in the context of local constant, kernel-based estimators of the regression mean (see Fan and Truong, 1993), although the methods can be generalised to other settings, for example to the estimators of Hall and Meister (2007) based on ridging or to the local polynomial estimators of Delaigle et al. (2009). In order to be fully practicable, the method requires choice of several bandwidths. Data-driven bandwidths have been developed in the errors-in-variables literature for the problem of regression estimation, but in order to ensure consistency of the coverage of the confidence bands, the bandwidths cannot all be taken to be of the size appropriate for estimating a regression curve. We introduce methods for bandwidth choice, focused specifically on the problem of ensuring good coverage accuracy. We also outline theoretical properties of our bootstrap bands.

We make several novel contributions: (1) We develop new strategies for generating the bootstrap sample in the context of errors-in-variables data, in a way that guarantees consistency of the bootstrap bands; see Sections 2.3 and 3.1. (2) We clarify the conditions that the smoothing parameters need to satisfy in order to guarantee consistency of the bands, and develop a novel and completely data-driven approach to selecting these parameters in a way that satisfies these conditions; see Section 3.2. (3) We develop theoretical properties of our bands; see Section 5.

2 Model and outline of methodology

2.1 Model and regression estimator

Suppose we observe data pairs (W_i, Y_i) , generated from the errors-in-variables regression model

$$W_i = X_i + U_i, \quad Y_i = g(X_i) + V_i \quad \text{for } 1 \leq i \leq n, \quad (2.1)$$

where the variables U_i, V_i, X_i for $1 \leq i \leq n$ are totally independent, the U_i s are identically distributed as $U \sim f_U$, the V_i s are identically distributed as $V \sim f_V$ and have zero mean and finite variance, and the X_i s are identically distributed as $X \sim f_X$. In this model, the explanatory variables X_i are not available, and we observe only their contaminated versions $W_i \sim f_W = f_X * f_U$, which are measured with unknown errors U_i . For simplicity, throughout most of this work we make the standard assumption that the error density f_U is known. The unknown error case will be treated in Section 4.2.

A popular nonparametric estimator of g , in the errors-in-variables model described at (2.1), is the local constant deconvolution estimator of Fan and Truong (1993). To define it, let K be a kernel function, write $\phi_K(t) = \int e^{itx} K(x) dx$ for its Fourier transform, let ϕ_U be the characteristic function of the distribution of U . We assume throughout that

$$\phi_U \text{ does not vanish on the real line,} \quad (2.2)$$

and that ϕ_K is compactly supported. Let $h > 0$ denote a bandwidth, and put

$$K_U(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu} \frac{\phi_K(t)}{\phi_U(t/h)} dt. \quad (2.3)$$

The local constant deconvolution estimator of g is defined by

$$\hat{g}(x) = \hat{g}(x; h) = \hat{a}(x) / \hat{f}_X(x), \quad (2.4)$$

where

$$\hat{f}_X(x) = \hat{f}_X(x; h) = \frac{1}{nh} \sum_{j=1}^n K_U\left(\frac{x - W_j}{h}\right) \quad (2.5)$$

is the deconvolution kernel estimator of f_X defined in Carroll and Hall (1988) and Stefanski and Carroll (1990), and where

$$\hat{a}(x) = \hat{a}(x; h) = \frac{1}{nh} \sum_{j=1}^n Y_j K_U\left(\frac{x - W_j}{h}\right) \quad (2.6)$$

denotes a kernel estimator of $a = f_X g$.

2.2 Pointwise confidence bands based on central limit theory

In standard error-free nonparametric curve estimation problems, a common method for constructing confidence bands is through the limiting distribution of estimators (see e.g. Härdle, 1989a, Eubank and Speckman, 1993, and Xia, 1998). In theory, such procedures could be employed in the errors-in-variables setting too. However, in practice, unlike the error-free case or density deconvolution problems, there does not seem to be an attractive way of estimating consistently the quantities required to construct the interval. See Section C.1 in the Supplementary Material, where we explain the difficulties associated with this approach. Therefore, while they can be defined, confidence bands based on limiting distributions are not really practicable in the errors-in-variables context.

2.3 Pointwise confidence bands based on the bootstrap

The difficulties raised in Section 2.2 make it unattractive to use confidence bands based on central limit results, or other procedures that rely explicitly on estimating the variance of \hat{g} . Instead we suggest constructing confidence bands using the bootstrap. As usual, we have a choice of percentile- t or percentile methods. However, using the percentile- t method requires us to estimate the variance of \hat{g} when bootstrapping the studentised form of $\hat{g} - g$, while percentile bootstrap approaches do not rely on studentisation. Moreover, when confining attention to two-sided confidence bands, the coverage accuracies of percentile and percentile- t confidence bands are of the same orders (see e.g. Sections 4.4 and 4.5 in Hall, 1992b). Therefore we shall use the percentile approach for bootstrap confidence bands.

Bootstrap procedures have been used in the error-free case; see, for example, Härdle

and Bowman (1988), Hall (1992c), Neumann and Polzehl (1998) and Claeskens and Van Keilegom (2003). However, the methods cannot simply be adapted to the error case, where the X_i s are not observed. In particular:

1. We need to develop new strategies for generating the bootstrap sample, that are valid for data contaminated with errors. This will be tackled below and in Section 3.1, where we shall develop a new moment-matching method for errors-in-variables data.
2. We need to develop a data-driven approach to selecting the smoothing parameters involved. This task is particularly challenging, since, even in the simple error-free case, the literature does not seem to indicate clearly how to do this completely adaptively. Perhaps surprisingly, we shall argue in Section 3.2 that choosing the tuning parameters is actually simpler in the errors-in-variables case than in the standard error-free case. We shall clearly derive conditions that the smoothing parameters should satisfy in order to guarantee consistency of the bands, and we suggest a novel approach especially designed for, and only valid in, the errors-in-variables setting.
3. We need to develop theoretical properties of our bands. These are considerably more difficult to obtain than in the error-free case, and will be handled in Section 5. As often in errors-in-variables problems, the proofs are quite technical. They will be relegated to the Supplementary Material.

To generate bootstrap resamples, let h_0 and h_1 be two bandwidths. Their choice and that of other bandwidths will be discussed in Section 3.2. Construct $\hat{g}(x; h_0)$ as in (2.4) using the bandwidth h_0 , and let $\mathcal{U}^* = \{U_i^*\}_{i=1}^n$, $\mathcal{X}^* = \{X_i^*\}_{i=1}^n$ and $\mathcal{V}^* = \{V_i^*\}_{i=1}^n$ denote resamples drawn by sampling randomly from the distributions with distribution functions F_U , $\tilde{F}_X(\cdot; h_1)$ and \tilde{F}_V , respectively, where $\tilde{F}_X(\cdot; h_1)$ is the estimator of F_X with bandwidth h_1 , defined in Hall and Lahiri (2008) (see Section D.2 in the Supplementary Material), and \tilde{F}_V is constructed by moment matching, see Section 3.1. The resamples are drawn in such a manner that, conditional on the original dataset $\mathcal{Z} = \{(W_1, Y_1), \dots, (W_n, Y_n)\}$, they are comprised of independent data and are independent of one another. The bootstrap version

of the model at (2.1) is:

$$W_i^* = X_i^* + U_i^*, \quad Y_i^* = \hat{g}(X_i^*; h_0) + V_i^* \quad \text{for } 1 \leq i \leq n. \quad (2.7)$$

Using the bootstrap data set $\mathcal{Z}^* = \{(W_i^*, Y_i^*)\}_{i=1}^n$ in place of the original data \mathcal{Z} , compute the bootstrap analogue \hat{g}^* of \hat{g} , replacing h by a bandwidth h_2 , chosen to mimic the distribution of $\hat{g}(x; h) - g(x)$ by that of $\hat{g}^*(x; h_2) - \hat{g}(x; h_0)$. In practice, $h_2 = h$ is an appropriate choice, but it is convenient in our discussion at this point to distinguish between the two bandwidths.

Let the nominal coverage of the confidence band be $1 - \alpha$, where $0 < \alpha < 1$. We define a nominal $(1 - \alpha)$ -level percentile pointwise confidence band for g , constructed over the compact interval \mathcal{I} , by:

$$\text{CB}_\alpha(\mathcal{I}) = \{(x, y) : x \in \mathcal{I} \text{ and } \hat{g}(x; h) - \hat{t}_{1-\alpha/2}(x) \leq y \leq \hat{g}(x; h) - \hat{t}_{\alpha/2}(x)\}, \quad (2.8)$$

where, for each x , $\hat{t}_{\alpha/2}(x)$ and $\hat{t}_{1-\alpha/2}(x)$ are such that

$$P\{\hat{g}^*(x; h_2) - \hat{g}(x; h_0) \leq \hat{t}_{\alpha/2}(x) | \mathcal{Z}\} = P\{\hat{g}^*(x; h_2) - \hat{g}(x; h_0) > \hat{t}_{1-\alpha/2}(x) | \mathcal{Z}\} = \frac{\alpha}{2}. \quad (2.9)$$

As is usually the case for bootstrap methods, the quantiles $\hat{t}_{\alpha/2}$ are computed as the empirical quantiles obtained from B bootstrap samples \mathcal{Z}_b^* , $b = 1, \dots, B$, generated according to (2.7). In practice, owing to the discreteness of both the bootstrap distribution and the Monte Carlo approximation algorithm, in most instances the identity “ $= \frac{\alpha}{2}$ ” can be achieved only approximately.

The pointwise nature of the equal-tailed band at (2.8) is reflected by the fact that it gives approximately correct coverage at individual points x , i.e.

$$P\{\hat{g}(x; h) - \hat{t}_{1-\alpha/2}(x) \leq g(x) \leq \hat{g}(x; h) - \hat{t}_{\alpha/2}(x)\} \rightarrow 1 - \alpha \quad \text{for all } x \in \mathcal{I}, \quad (2.10)$$

as sample size diverges. See Section 5.2.

Remark 1. The band defined by (2.8) is the one we employed in our numerical work, but other $(1 - \alpha)$ -level bands for g can also be defined. See for example Section D.1 in the Supplementary Material for an alternative equal-tailed band and a symmetric band.

2.4 Simultaneous confidence bands

In this work, we focus on pointwise bands. However, it is worth mentioning that our bootstrap methodology would be largely unchanged if the bands were simultaneous; the main alteration would be to the critical point determining interval width. More precisely, a simultaneous analogue of the pointwise confidence band defined in (2.8) can be constructed by altering equation (2.9) to

$$\begin{aligned} P\{\hat{g}^*(x; h_2) - \hat{g}(x; h_0) \leq \hat{t}_{\alpha/2}(x) \text{ for all } x \in \mathcal{I} | \mathcal{Z}\} \\ = P\{\hat{g}^*(x; h_2) - \hat{g}(x; h_0) > \hat{t}_{1-\alpha/2}(x) \text{ for all } x \in \mathcal{I} | \mathcal{Z}\} = \frac{\alpha}{2}. \end{aligned}$$

In this case, and modulo appropriate assumptions, the following approximation would prevail in place of (2.10): $P\{\hat{g}(x; h) - \hat{t}_{1-\alpha/2}(x) \leq g(x) \leq \hat{g}(x; h) - \hat{t}_{\alpha/2}(x) \text{ for all } x \in \mathcal{I}\} \rightarrow 1 - \alpha$.

Simultaneous bands based on asymptotic arguments can also be considered. They are generally founded on approximations to the distributions of Gaussian processes or sequences of normally distributed random variables (see e.g. Bickel and Rosenblatt, 1973). In consequence they are afflicted by the inferior accuracy of such approximations (see e.g. Hall, 1979), and can have very poor performance. Therefore, asymptotic methods are even less attractive than in the pointwise case discussed in Section 2.2, and bootstrap methods give more satisfactory results. For example, in the so-called ordinary smooth case, where the characteristic function of U decays polynomially fast in the tails, the difference, in the setting of simultaneous confidence intervals, is that coverage errors of size $(\log n)^{-1/2}$ occur when using asymptotic methods, whereas accuracy is of order n^{-c} , for some $c > 0$, in the bootstrap case. See e.g. Hall (1991). Theory, too, is quite different in pointwise and simultaneous cases.

3 Details of percentile bootstrap methodology

3.1 Estimating the distribution of V

The distribution function F_V of V is not readily accessible from data of the type at (2.1), because we know neither the function g nor the variables X_i . Since $F_Y = f_R * F_V$, where

F_Y is the distribution function of Y and f_R is the density of $R = g(X)$, both of which can be estimated from the data, we could estimate F_V consistently using a nonparametric regularised deconvolution procedure. This would provide a consistent estimator of F_V , but it is unattractive, not least because it suffers from slow convergence rates.

Fortunately, the main properties of confidence bands for g depend on the distribution of V only through its first three moments (see Section 5.1 for justification), and these are relatively easy to access. Therefore, it suffices to generate our resampled data V_i^* from a distribution whose first three moments are consistent with those of F_V . These requirements characterise moment-matching bootstrap methods, which date from a suggestion by Wu (1986). See also Liu (1988), Cao-Abad (1991) and Mammen (1992).

The difficulty in our errors-in-variables context is to construct consistent moment estimators. Since we know that $E(V) = 0$, we put the first moment estimator equal to zero. We construct consistent estimators of $\sigma^2 = E(V^2)$ and $\zeta = E(V^3)$ under the assumption that $E(V^6) + E\{g(X)^6\} < \infty$. Let $\mu_j = E(Y^j)$ and $\xi_j = E\{g(X)^j\}$. Since V and X are independent and $E(V) = 0$, it follows from (2.1) that $\mu_2 = \xi_2 + \sigma^2$ and $\mu_3 = \xi_3 + 3\xi_1\sigma^2 + \zeta$. We suggest estimating μ_j and ξ_j by $\hat{\mu}_j = n^{-1} \sum_i Y_i^j$ and $\hat{\xi}_j = \int \hat{g}(x)^j d\tilde{F}_X(x) = E\{\hat{g}(X^*)^j | \mathcal{Z}\}$, respectively. (In practice, we compute the value of $\hat{\xi}_j$ as the empirical average of many $\hat{g}(X_i^*)$ s, where each X_i^* is drawn from \tilde{F}_X .) Finally, our estimators of σ^2 and ζ are

$$\hat{\sigma}^2 = \max(\hat{\mu}_2 - \hat{\xi}_2, 0), \quad \hat{\zeta} = \hat{\mu}_3 - \hat{\xi}_3 - 3\hat{\xi}_1\hat{\sigma}^2, \quad (3.1)$$

respectively. See Section D.3 of the Supplementary Material for a discussion of consistency of $\hat{\sigma}^2$ and $\hat{\zeta}$. See also Delaigle and Hall (2011) for an alternative, more complex estimator of σ^2 .

We can generate V_i^* from any distribution with first three moments equal to zero, $\hat{\sigma}^2$ and $\hat{\zeta}$, respectively. In our numerical work we used a centred, rescaled gamma distribution, with density

$$f(x) = \frac{(x + k\theta)^{k-1}}{\theta^k \Gamma(k)} e^{-(x+k\theta)/\theta} \cdot 1\{x > k\theta\}, \quad (3.2)$$

where $k, \theta > 0$. This density produces a distribution with positive third moment, and we fold the distribution into the negative half-line if $\hat{\zeta} < 0$; when $\hat{\zeta} = 0$ this procedure produces

the $N(0, \hat{\sigma}^2)$ model for the distribution of V .

3.2 Choice of bandwidths

To construct our bootstrap confidence band at (2.8), we need to select four bandwidths: h , h_0 , h_1 and h_2 . As we shall see in Section 5.2, if all four bandwidths are of conventional size for estimating g , then the effect of the bias of the regression estimator on the coverage error of bootstrap confidence bands can be significant, and needs to be accommodated in some way. This is well known in the standard error-free case; see e.g. Härdle and Bowman (1988), Härdle (1989a), Härdle and Marron (1991), Hall (1992a), Neumann (1995), Neumann and Polzehl (1998) and Cummins et al. (2001).

In principle we could do explicit bias correction, but this is an unattractive approach due to the complexity of bias estimators in errors-in-variables problems, as mentioned in Section 2.2 (see also Section C.1 in the Supplementary Material). Instead, in Section 5.2, we shall derive two different sets of conditions on the orders of magnitude of the bandwidths that each ensure consistency of the bands: (1) take h , h_1 and h_2 of conventional size, and take h_0 larger than usual; or (2) take h_0 and h_1 of standard size, and h and h_2 smaller than usual. Option (2) is not easy to implement in practice, even in the error-free case, because it requires choosing two non conventional bandwidths in an automatic data-driven way, and it is not clear how to do this in practice. Moreover, while undersmoothing is necessary to obtain consistent bands, too much undersmoothing can produce too wiggly bands. For all these reasons, we use the first approach.

In more details, we take h_1 equal to the normal reference bandwidth for deconvolution density estimation described in Delaigle and Gijbels (2004), and we take $h = h_2 = \hat{h}$, found by the SIMEX (SIMulation EXtrapolation) regression bandwidth selection procedure of Delaigle and Hall (2008). For h_0 , while we know that it has to be larger than usual, it remains to find a way of choosing it in practice. Even in the error-free case there is no easy way to do this, but, ironically, we shall see below that the errors-in-variables structure has an advantage because it enables us to estimate coverage error, and we shall exploit this fact to suggest a method for choosing h_0 .

First, we define the ideal, theoretical, h_0 to be the bandwidth that minimizes a global measure of coverage error over an interval \mathcal{J} . More precisely, we take $h_0 = c_0 h$, where

$$c_0 = \operatorname{argmin}_{c \geq 1} \int_{\mathcal{J}} \{\operatorname{CP}(x; c) - (1 - \alpha)\}^2 dx, \quad (3.3)$$

with $\operatorname{CP}(x; c) = P\{g(x) \in \operatorname{CB}_{ch}(x)\}$, and $\operatorname{CB}_{ch}(x)$ denoting the band defined at (2.8), replacing there h_0 by ch . In Section D.4 of the Supplementary Material, we shall prove that h_0 defined in this way is an order of magnitude larger than h , h_1 and h_2 , and thus guarantees consistency of the band.

Of course we do not know g , thus we cannot calculate CP. To estimate it, we develop a SIMEX procedure for this problem. The SIMEX method was introduced by Cook and Stefanski (1994) for parametric estimation of curves in errors-in-variables problems. It involves generating new data, with further noise added to existing measurement errors on the X_i s. Since we contaminated these new data artificially, we have access to their error-free version (the version before we added further noise), as well as their contaminated version (the version after we added the extra noise). Therefore we can learn about the effect that measurement errors has on a target, and extrapolate back this effect to the original dataset.

In the nonparametric context, Delaigle and Hall (2008) suggested a SIMEX method for selecting bandwidth for estimating a regression curve, but it cannot be used to select h_0 , since it provides bandwidths of the same magnitude as h and h_2 . In our confidence band context, we suggest the following new SIMEX bandwidth selection procedure, that aims to mimick the behaviour of $h_0 = c_0 h$, with c_0 as in (3.3). Theoretical arguments will be provided in Section D.4 in the Supplementary Material.

First, generate the SIMEX data $W_{1,i}$ and $W_{2,i}$, for $1 \leq i \leq n$, in the following way:

1. Generate $\{U_{1,i}\}_{i=1}^n$ and $\{U_{2,i}\}_{i=1}^n$ from the known distribution of U .
2. Put $W_{1,i} = W_i + U_{1,i}$ and $W_{2,i} = W_{1,i} + U_{2,i}$ for $1 \leq i \leq n$.

Next, the idea is to regard temporarily the problem as one of constructing confidence bands for the regression of Y_i on W_i and of Y_i on $W_{1,i}$, rather than of Y_i on X_i . For this, define $g^+(x) = E(Y_i | W_i = x)$ and $g^{++}(x) = E(Y_i | W_{1,i} = x)$. The samples $\mathcal{Z}^+ = \{(W_{1,i}, Y_i)\}_{i=1}^n$

and $\mathcal{Z}^{++} = \{(W_{2,i}, Y_i)\}_{i=1}^n$ are, respectively, versions of $\mathcal{Z} = \{(W_i, Y_i)\}_{i=1}^n$ and $\mathcal{Z}^+ = \{(W_{1,i}, Y_i)\}_{i=1}^n$ contaminated by errors with density f_U . For $\omega = h, \hat{h}, h_0, h_j$ and $\hat{h}_j, j = 1, 2$, let ω^+ and ω^{++} denote, respectively, the version of ω calculated as above, but replacing \mathcal{Z} by \mathcal{Z}^+ and \mathcal{Z}^{++} . Likewise, let $c_0^+, \text{CB}_{h_0}^+$ and CP^+ be defined as c_0, CB_{h_0} and CP above, but replacing \mathcal{Z} by \mathcal{Z}^+ , g by g^+ and each bandwidth ω by ω^+ . In particular, $\text{CB}_{h_0}^+$ is a confidence band for g^+ and $\text{CP}^+(x; c) = P\{g^+(x) \in \text{CB}_{ch^+}^+(x)\}$. Define $\omega^{++}, c_0^{++}, \text{CB}_{h_0}^{++}$ and CP^{++} similarly, this time replacing \mathcal{Z} by \mathcal{Z}^{++} , g by g^{++} and each bandwidth ω by ω^{++} . In particular, $\text{CB}_{h_0}^{++}$ is a confidence band for g^{++} and $\text{CP}^{++}(x; c) = P\{g^{++}(x) \in \text{CB}_{ch^{++}}^{++}(x)\}$.

Of course, g^+ and g^{++} are unknown, but since we have access to direct samples from (W_i, Y) and $(W_{1,i}, Y)$, we can construct standard error-free nonparametric estimators \hat{g}_{EF}^+ and \hat{g}_{EF}^{++} , which converge to g^+ and g^{++} at much faster rates than in errors-in-variables problems. (By error-free estimators, we mean estimators based on non-contaminated data, such as the standard local linear estimator.) In particular, \hat{g}_{EF}^+ and \hat{g}_{EF}^{++} have negligible estimation error compared to the estimation of error of CB^+ and CB^{++} , which are constructed from contaminated data. Motivated by this, we repeat the procedure S times, for each of S SIMEX samples, obtaining estimates $\hat{g}_{s,EF}^+, \hat{g}_{s,EF}^{++}, s = 1, \dots, S$ and confidence bands $\text{CB}_{s, ch_s^+}^+$ and $\text{CB}_{s, ch_s^{++}}^{++}$, for the s th SIMEX sample, where $s = 1, \dots, S$. We suggest estimating the coverage probabilities for $g^+(x)$ and $g^{++}(x)$ by

$$\widehat{\text{CP}}^+(x; c) = \frac{1}{S} \sum_{s=1}^S I\{\hat{g}_{s,EF}^+(x) \in \text{CB}_{s, ch_s^+}^+(x)\},$$

$$\widehat{\text{CP}}^{++}(x; c) = \frac{1}{S} \sum_{s=1}^S I\{\hat{g}_{s,EF}^{++}(x) \in \text{CB}_{s, ch_s^{++}}^{++}(x)\},$$

and we take

$$\hat{c}_0^+ = \operatorname{argmin}_{c \geq 1} \int_{\mathcal{J}} \{\widehat{\text{CP}}^+(x; c) - (1 - \alpha)\}^2 dx.$$

We define c_0^{++} in the same way, replacing $^+$ by $^{++}$.

Then, by using reverse extrapolation, we can compute an approximation to the factor c_0 for the original problem. More precisely, since CB, CB^+ and CB^{++} are all constructed from data contaminated by errors with density f_U , then c_0, c_0^+ and c_0^{++} all have the same asymptotic order as h_0/h . Further, quoting Delaigle and Hall (2008) (although replacing

there the bandwidths by the factors c_0 , c_0^+ and c_0^{++}), since, for $k = 0, 1$, $W_{k+1,i}$ measures $W_{k,i}$ in the same way as W_i measures X_i , it can be expected that c_0^{++} measures c_0^+ in the same way as c_0^+ measures c_0 , that is, $c_0^{++}/c_0^+ \approx c_0^+/c_0$, so that $c_0 \approx (c_0^+)^2/c_0^{++}$. Therefore, we estimate c_0 by $\hat{c}_0 = (\hat{c}_0^+)^2/\hat{c}_0^{++}$, and take $\hat{h}_0 = \hat{c}_0\hat{h}$. See Section D.4 of the Supplementary Material for theoretical arguments justifying our approach.

Remark 2. The SIMEX method could even be used to construct the entire confidence band, as follows. First, construct standard error-free confidence bands for g^+ , g^{++} , g^{+++} , where $g^{+++}(x) = E(Y_i|W_{2,i} = x)$, using, respectively, the samples \mathcal{Z} , \mathcal{Z}^+ and \mathcal{Z}^{++} . Then extrapolate these bands, to obtain SIMEX-based confidence bands for g . This method, or refined versions of it, are straightforward to implement. However, just as SIMEX does not, in general, lead to consistent estimators of regression means in errors-in-variables problems, the SIMEX confidence bands constructed in this way do not have asymptotically correct coverage, in contrast with bands based on the bootstrap.

3.3 Calibrating the confidence bands

To improve coverage accuracy of our confidence bands in practice, we calibrate them through a standard double bootstrap method, as follows (see for example Beran, 1987, Loh, 1987, Hall and Martin, 1988 and Martin, 1990). Let $\pi_0(\alpha)$ denote the actual coverage of the band CB_α constructed from the original sample \mathcal{Z} , that is, $P\{g(x) \in CB_\alpha(x)\} = \pi_0(\alpha)$. Instead of constructing CB_α , if we construct CB_{β_α} where β_α is such that $\pi_0(\beta_\alpha) = 1 - \alpha$, then the band CB_{β_α} will have the nominal coverage $1 - \alpha$. As in more standard problems, we estimate β_α by solving $P\{\hat{g}(x) \in CB_{\hat{\beta}_\alpha}^*(x)|\mathcal{Z}\} = 1 - \alpha$, where $CB_{\hat{\beta}_\alpha}^*$ denotes a band computed defined in Section 2.3, but replacing there the data \mathcal{Z} by a bootstrap sample \mathcal{Z}^* . Of course, this involves double bootstrap since, computing $CB_{\hat{\beta}_\alpha}^*$ requires drawing bootstrap samples, say C , from the bootstrap sample \mathcal{Z}^* . See Appendix B for details.

4 Numerical properties

4.1 Details of implementation

We applied our procedure for constructing the percentile bootstrap confidence band described in Section 2.3 to real and simulated data, using the bandwidths described in Section 3. In each case we took the interval \mathcal{J} where the integrals in Section 3 are calculated, equal to $[q_{0.1}, q_{0.9}]$, $[q_{0.1}^+, q_{0.9}^+]$ and $[q_{0.1}^{++}, q_{0.9}^{++}]$, respectively when calculations were based on the sample \mathcal{Z} , \mathcal{Z}^+ and \mathcal{Z}^{++} , respectively, and where q_p , q_p^+ and q_p^{++} denote, respectively, the empirical p th quantile of the W_{iS} s, the $W_{1,iS}$ s and the $W_{2,iS}$ s (the latter two quantiles are calculated from one SIMEX sample). For the \mathcal{J} used in Section 3.3, we took $\mathcal{J} = [q_{0.1}, q_{0.9}]$, where q_p is defined above. We took the number of bootstrap and double bootstrap replications equal to, respectively, $B = 100$ and $C = 100$. The SIMEX calculations were based on $S = 20$ SIMEX samples.

The standard error-free regression estimators in the SIMEX methodology were calculated using a local linear estimator with a standard normal kernel and a cross-validation bandwidth. The errors-in-variables estimators were calculated using the kernel K , defined through its characteristic function by $\phi_K(t) = (1 - t^2)^3 \cdot 1\{t \in [-1, 1]\}$. To gain time, when calculating the various bandwidths involved, we binned the data in $[n/5]$ bins; see Appendix D in Wand and Jones (1995).

4.2 Unknown error case

In practice, the error density f_U is not always known. As usual in errors-in-variables problems, our procedure can be adapted to this case as long as we have repeated measurements

$$W_{jk} = X_j + U_{jk} \quad \text{for } k = 1, 2,$$

and the U_{jk} s are distributed as U and are independent of the X_j s. Let $\bar{W}_j = (W_{j1} + W_{j2})/2 = X_j + \bar{U}_j$ with $\bar{U}_j = (U_{j1} + U_{j2})/2$. We apply our procedure with the sample $\{(\bar{W}_j, Y_j)\}_{j=1}^n$. In particular, here the errors are \bar{U}_j , and their characteristic function is $\phi_{\bar{U}}(t) = E(e^{it\bar{U}_j})$.

To construct bootstrap and SIMEX samples, we need to generate data from the distribution of \bar{U}_i , which is unknown. We construct the sample $\bar{U}_1^*, \dots, \bar{U}_n^*$ by drawing with replacement from the data set $\{(W_{j1} - W_{j2})/2\}_{j=1}^n$. We proceed similarly to generate the errors required to compute the SIMEX samples, as suggested by Delaigle and Hall (2008). This approach to generating new samples of error variables is particularly simple; for example it does not require any additional bandwidth.

Next, to compute \hat{f}_X and \hat{g} , we need to derive an estimator of the unknown $\phi_{\bar{U}}$. Since the errors \bar{U}_i have the same distribution as $(W_{i1} - W_{i2})/2$, as in Delaigle et al. (2008), we can estimate $\phi_{\bar{U}}(t)$ by $\hat{\phi}_{\bar{U}}(t) = n^{-1} \sum_{j=1}^n \cos \{t(W_{j1} - W_{j2})/2\}$.

4.3 Simulations

We calculated our bootstrap confidence bands from samples of size $n = 100$ or $n = 200$ generated from the model at (2.1), with each of the following regression functions: $g_1(x) = 1.5 \cos\{(x+3)/2\}$, $g_2(x) = \{(x-1)^2/17\} - 1$, $g_3(x) = (x/4) + 4(2\pi)^{-1/2} e^{-x^2/8}$, and $g_4(x) = 1.5 \sin\{(x-1)/2.3\}$. In each case, we took $X_i \sim N(0, 8)$ and $V_i \sim N(0, 0.04)$, while the U_i s were generated according to distributions usually encountered in the errors-in-variables literature, namely, a Laplace distribution with parameter σ , $\text{Lap}(\sigma)$, the convolution of a $\text{Lap}(\sigma)$ distribution with itself, $(\text{Lap} * \text{Lap})(\sigma)$, or a normal distribution. More precisely, they were generated according to one of the following distributions: E_A : $\text{Lap}(2^{-1/2})$; E_B : $\text{Lap}(\sqrt{0.8})$; E_C : $\text{Lap} * \text{Lap}(0.5)$, E_D : $(\text{Lap} * \text{Lap})(0.6)$ or E_E : $N(0, 1)$. These particular curves and distributions were chosen so that the graph of g , for a typical sample of size $n = 100$ or $n = 200$, contains features such as a peak and/or a valley, and the noise-to-signal ratio (NSR), $\text{Var}(U)/\text{Var}(X)$ was approximately equal to $1/8$ (cases E_A , E_C and E_E) or $1/5$ (cases E_B and E_D).

In each case, we generated 100 samples from one of the settings described above, and for each sample we calculated the 95% bootstrap confidence band obtained using the algorithm described in Section 2.3 for each x on an equispaced grid in $\in [-5, 5]$. Regarding the error distribution F_U , we considered three cases: F_U is known (KNE); F_U is unknown and, as in Section 4.2, is estimated through replicated measurements $W_{ij} = X_i + U_{ij}$, $j = 1, 2$

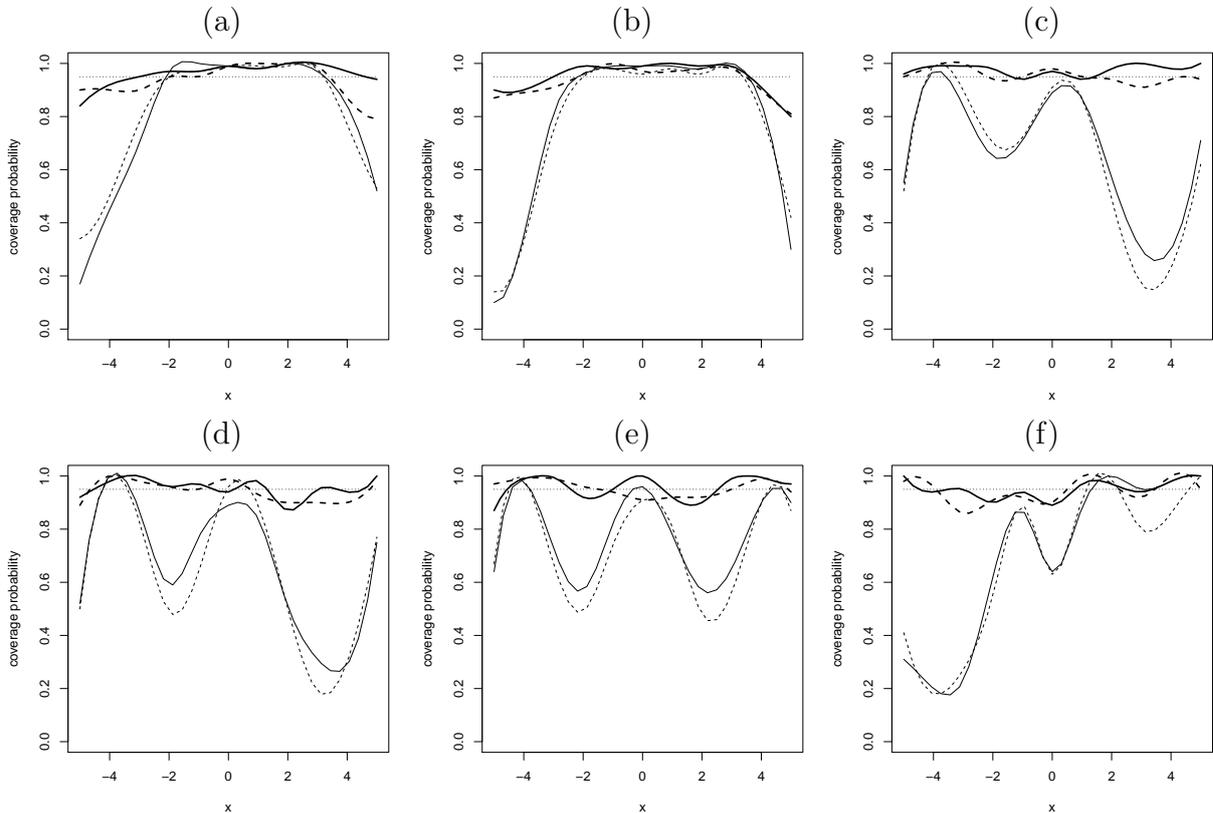


Figure 1: Coverage probability curves derived from our bootstrap method in the KNE case with $n = 100$ (thick dashed line) and $n = 200$ (thick solid line), and from the naive approach with $n = 100$ (thin dashed line) and $n = 200$ (thin solid line). (a): $g = g_2$ and error E_A , (b): $g = g_2$ and error E_B , (c): $g = g_4$ and error E_B , (d): $g = g_4$ and error E_D , (e): $g = g_1$ and error E_C and (f): $g = g_3$ and error E_E . The horizontal dotted line indicates 0.95 for reference.

(EST), where the U_{ij} s are i.i.d. and independent of the X_i s; and F_U is unknown and the bands are computed as if the errors were Laplace (LAP). The first setting is the most commonly assumed in the nonparametric deconvolution literature. The second setting is often encountered in practice but has only recently been considered in the nonparametric literature; see Li and Vuong (1998) for early development, and Delaigle et al. (2008) for a practicable approach. Note that in this setting, we took the U_{ij} s to have the same distribution as the U_i s in E_A – E_E above, except that we took the scale parameter so that $\text{var}(U_{ij}) = 2\text{var}(U_i)$. The third setting is introduced here only to study the robustness of the bands against error misspecification, as requested by an associate editor. Note that as argued by Meister (2004) and Delaigle (2008), when the error distribution is unknown, it is preferable to assume that it is Laplace, whence our choice of the Laplace distribution.

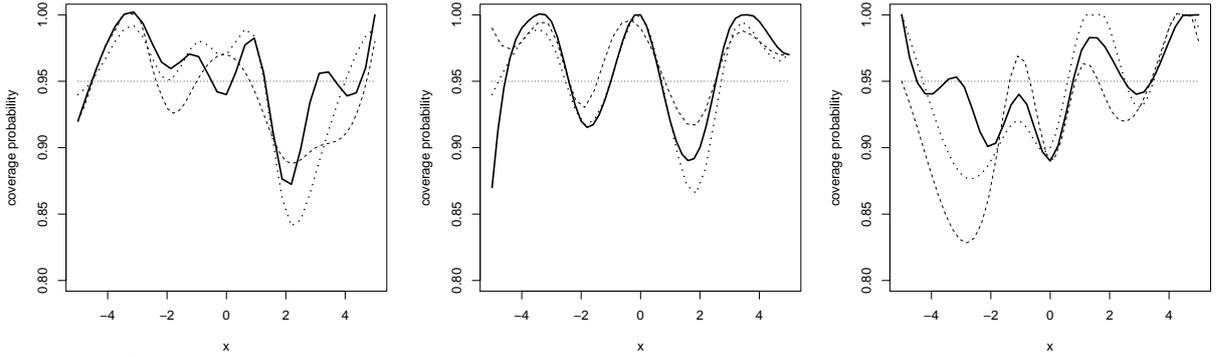


Figure 2: Coverage probability curves derived from our bootstrap method in the KNE case (thick solid line), the LAP case (dotted line) and the EST case (dashed line) with $n = 200$. Left: $g = g_4$ and error E_D ; middle: $g = g_1$ and error E_C ; right: $g = g_3$ and error E_E . The horizontal dotted line indicates 0.95 for reference.

To illustrate the importance of taking the measurement error into account when constructing confidence bands, we also computed 95% confidence bands derived from the so-called naive estimator, i.e. a standard nonparametric regression estimator (e.g. Nadaraya-Watson) constructed from the data (W_i, Y_i) , pretending they were observed without measurement errors. To construct these naive bands, we cannot adapt our method to the error-free case, since it relies crucially on SIMEX, which can only be calculated for errors-in-variables estimators. Instead we used bands for standard nonparametric estimators that are available in R, namely those based on asymptotic normal distribution, which can be calculated using the package `locfit`.

We present the results of several examples, selected to illustrate the properties of our method in different settings. We obtained similar results in other settings. First, in Figure 1 we show graphs of estimated coverage probabilities for $x \in [-5, 5]$, calculated in each case from 100 samples and derived using our bootstrap method in the KNE case, and the naive approach, for samples of size $n = 100$ and 200, and for various combinations of curves and error distributions. We can see that, despite the difficulty of the problem, when using our bootstrap approach, for most $x \in [-5, 5]$ the coverage probabilities are close to the nominal level 0.95 when $n = 200$, although as usual the coverage is poorer at x -values closer to the tail of f_X ; as expected, the coverage improves as n increases from 100 to 200. By contrast, overall, the bands constructed by the naive approach that ignores the error have very poor

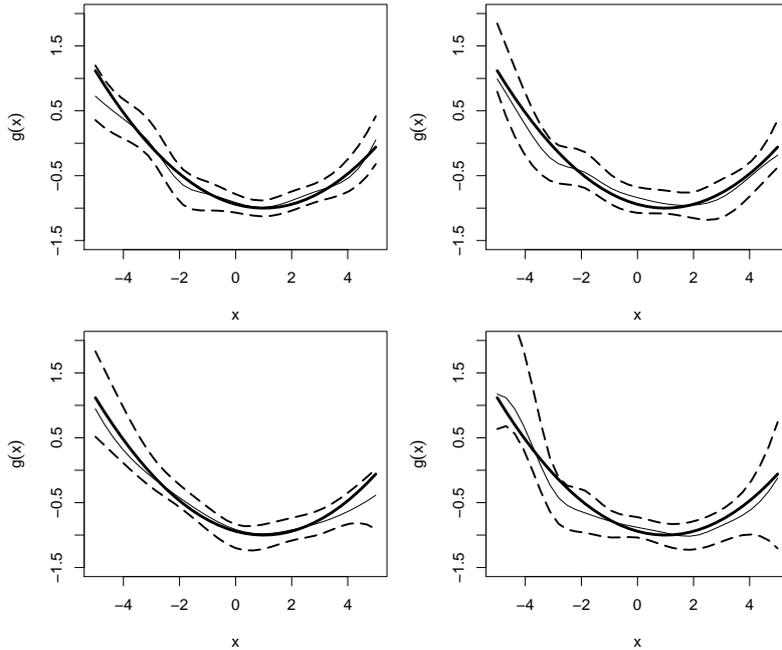


Figure 3: \hat{g} (thin solid line), $g = g_2$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 40 (left) and 60 (right), in the KNE case, when $n = 200$ and the error distribution is E_A (row 1) or E_B (row 2).

coverage at many x -values.

Figure 2 compares the coverage probability curves, for cases (d)–(e) of Figure 1 and when $n = 200$, obtained in the KNE, LAP and EST cases. Graphs for the case $n = 100$ are shown in Figure F1 in Appendix F in the Supplementary Material. These figures illustrate that, even when the error distribution is unknown and either misspecified to be Laplace or estimated from replicates, our procedure gives bands with good coverage, especially compared with naive bands that do not take measurement errors into account. In general, and unsurprisingly, the best coverage is obtained in the KNE case, although in the Gaussian error case, pretending that the error distribution is Laplace can sometimes improve coverage compare to the KNE case. This is not uncommon in deconvolution problems, and is due to the numerical instability of the Gaussian error case.

Next, we depict confidence bands constructed by our bootstrap method for several contaminated samples, selected as follows. For each case considered in our simulations, we generated 100 samples, and constructed the corresponding 100 estimators \hat{g} as in (2.4) and

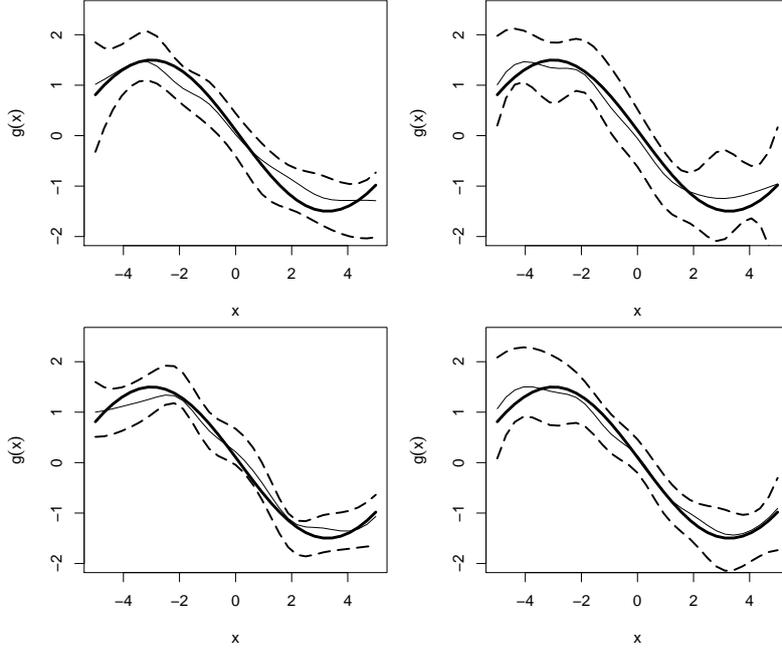


Figure 4: \hat{g} (thin solid line), $g = g_1$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 40 (left) and 60 (right), in the KNE case, when $n = 100$ (row 1) or 200 (row 2) and the error distribution is E_C .

their respective bootstrap bands. To choose which samples to show on the graphs, we ranked the samples according to a score based on coverage and width of the bands, as follows. In each case, the 100 bands were first ranked according to their overall coverage on the interval $[-5, 5]$, the highest coverage receiving the lowest rank. Then, to break ties, samples with the same coverage were ranked according to the area comprised within the band, the smallest area receiving the smallest rank. Proceeding this way, the samples received ranks 1 to 100. In each case, in the graphs, we show the confidence bands corresponding to the samples that received rank 40 and 60 in this ranking. This ranking permits us to compare bands obtained in different settings. Here we show the graphs corresponding to a selected set of examples illustrating the main properties of our method; for more examples, see Appendix F in the Supplementary Material.

Figure 3 shows the 95% confidence bands for $g = g_2$, when $n = 200$ and the error is Laplace with $\text{NSR} = 1/8$ or $\text{NSR} = 1/5$, in the KNE case. We can see that the bands are rather narrow, and, as could be expected, are narrower when the error is smaller. In Figure 4 we show, in the KNE case, the confidence bands for $g = g_1$ when the error distribution is

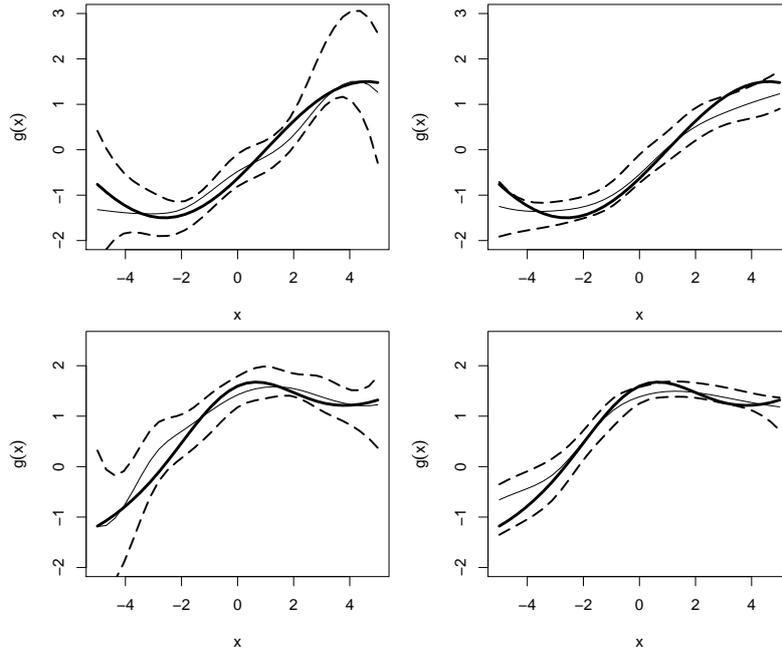


Figure 5: \hat{g} (thin solid line), $g = g_4$ (thick solid line) and 95% confidence bands (dashed lines) for g constructed from the samples ranked 40 (left) and 60 (right), in the EST case, when $n = 100$ and the error distribution is E_D (row 1) or E_E (row 2).

a convolution of two Laplace distributions, and $n = 100$ or $n = 200$. We see that the bands become more narrow as n increases. Figure 5 considers two challenging settings, in the first because sample size is small and error variance is large, and the second because the error is normal, which is notoriously difficult to deal with in errors-in-variables problems. More precisely, the figure shows, in the EST case, the confidence bands for $g = g_3$ when $n = 100$ and the error is a convolution of two Laplace distributions with $\text{NSR} = 20\%$, and $g = g_4$ when the error is normal. In those complex settings too we obtain reasonable bands, even though the error distribution is estimated from replicates.

Finally, in Figures F5–F10 in Appendix F in the Supplementary Material, to further illustrate the effect of estimating or misspecifying the error, we depict, for various combinations of regression curves g and error distributions, the confidence bands obtained in cases KNE, EST and LAP. The figure shows that we obtain reasonable confidence bands in all three cases, although quite logically, the case KNE generally gives the most attractive bands.

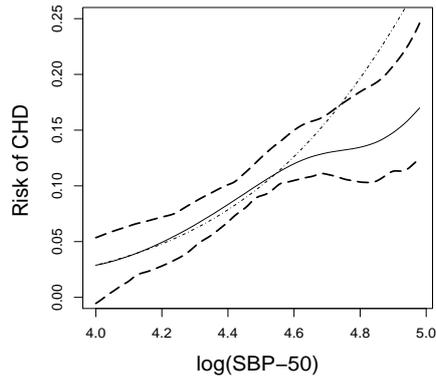


Figure 6: 95% confidence band for the risk of CHD; the dashed lines represent the confidence band limits, the solid line shows the nonparametric estimate of the risk of CHD, and the dot-dashed lines shows the logistic estimate.

4.4 Real Data Example

The suggested method of constructing confidence bands was applied to data from the Framingham heart study described Carroll et al. (2006). The data comprise measurements of systolic blood pressure (SBP) for 1,615 male patients, taken twice in each of two exams. As in Delaigle and Hall (2008), for $i = 1, \dots, 1615$, $j = 1, 2$, we let $W_{ij} = \log\{(\text{SBP}_{ij1} + \text{SBP}_{ij2})/2 - 50\}$, where SBP_{ij1} and SBP_{ij2} are the first and second measurements at exam j , respectively. For each patient, the variable Y_i , $i = 1, \dots, 1615$, indicating the presence ($Y_i = 1$) or absence ($Y_i = 0$) of coronary heart disease (CHD) over an 8-year follow-up period, is also available. Our interest is in the relationship between the risk of CHD and the SBP level. Here since the error distribution is unknown, we employed the procedure described in Section 4.2 to construct our confidence band.

Figure 6 shows the estimated regression curve, calculated using the SIMEX bandwidth of Delaigle and Hall (2008), and the 95% confidence band limits, constructed using our procedure. The figure also depicts the parametric logistic fit obtained by regression calibration (see e.g. Carroll et al., 2006). The logistic model seems reasonable for x in $[4, 4.7]$, and in that range, it does not differ much from the nonparametric regression estimator. However, for $x > 4.7$, the parametric fit does not lie in our confidence band, suggesting that it may be preferable to use a more sophisticated model. The nonparametric estimator seems to be a good alternative. As could be expected, the figure suggests that the risk of CHD increases

with the level of SBP.

5 Theoretical properties

5.1 Edgeworth and Cornish-Fisher expansions

In deconvolution problems, it is standard to consider two types of error distributions F_U , called ordinary smooth and supersmooth, respectively; see Fan (1991). The distribution F_U is ordinary smooth (respectively, supersmooth) if ϕ_U decays polynomially fast (respectively, exponentially fast) in the tails. Similarly to Bissantz et al. (2007), we derive our theoretical results for the ordinary smooth case, as our conditions will imply. Analogous results can be derived for the supersmooth case, but at the expense of longer arguments. The results of our numerical work (see Section 4) suggest that our procedure works reasonably well in the supersmooth error case too.

In conventional semiparametric problems, Edgeworth expansions of distribution functions have the form

$$P\{n^{1/2}(\hat{\theta} - \theta) \leq \tau w\} = \Phi(w) + n^{-1/2}p_1(w)\phi(w) + n^{-1}p_2(w)\phi(w) + O(n^{-3/2}). \quad (5.1)$$

Here, $\hat{\theta}$ is an estimator of a parameter θ , τ^2/n equals the asymptotic variance of $\hat{\theta}$, Φ and ϕ denote the standard normal distribution and density functions, respectively, and each p_j is a polynomial, being an odd and even function according as j is even or odd, respectively.

A close analogue of (5.1) holds in many nonparametric problems, the main change being to replace θ in (5.1) by either $E(\hat{\theta})$ or a close analogue of it, and to replace n , for example in the factors $n^{-1/2}, n^{-1}, \dots$, by a quantity that is of the same size as the inverse of the variance of $\hat{\theta}$. The problem is a little more complex even in the case of standard error-free nonparametric regression, since then a portion of the variance of $\hat{\theta}$ can be estimated root- n consistently while another part can be estimated only at the root-mean-square rate of $(nh)^{-1/2}$. This means that Edgeworth expansions in the case of nonparametric regression have a form which derives from competing stochastic errors of sizes $(nh)^{-1/2}$ and $n^{-1/2}$.

A similar phenomenon is observed in errors-in-variables regression. Here the analogue

of $\hat{\theta}$ is $\hat{g} = \hat{a}/\hat{f}_X$, and $g_1 = E(\hat{a})/E(\hat{f}_X)$ is a natural choice for θ . Moreover, the analogue of τ^2/n is the asymptotic variance of $\hat{g}(x; h)$, given by $\delta^2(h)A_0^2(x)$, where

$$\delta^2(h) = s^2(h)/\{E(\hat{f}_X)\}^2, \quad (5.2)$$

$$s^2(h) = \frac{1}{nh} \int K_U^2(x) dx = \frac{1}{2\pi nh} \int |\phi_K(t)|^2 |\phi_U(t/h)|^{-2} dt, \quad (5.3)$$

$$A_0^2(x) = \int [\{g(q) - g_1(x)\}^2 + \sigma^2] L_0^2(u) f_U(x - q - hu) f_X(q) dq du, \quad (5.4)$$

with $L_0 = K_U/(\int K_U^2)^{1/2}$. (Note that $\int K_U^2$ diverges as h decreases to zero.) This means that the analogue of the left-hand side of (5.1) is

$$P_1(w) = P[\delta(h)^{-1}\{\hat{g}(x; h) - g_1(x)\} \leq A_0 w].$$

This motivates Theorem 5.1 and Corollary 5.1 below, which give, respectively, Edgeworth and Cornish-Fisher expansions relating to the distribution of $\hat{g} - g_1$. Recall that \mathcal{I} denotes the compact interval over which the confidence band is constructed. The expansions will be established under Conditions A, B and C given in Appendix A. In particular, we develop results that hold uniformly in $x \in \mathcal{I}$, and our assumptions reflect this goal. In fact, Theorem 5.1 describes an expansion of the distribution of $\hat{g}(x)$ that is valid not just uniformly in x but also in a large class of functions g and distributions of V and X . This makes our regularity conditions a little cumbersome, but means that the result yields expansions of bootstrap distributions, where g and the distributions of V and X are actually random quantities. See the discussion below Corollary 5.1.

Before we state the results formally, we need to introduce notation. Recall the definitions of \hat{a} and \hat{f}_X at (2.6) and (2.5), and put

$$A_1(x) = \int L_0^2(u) f_U(x - q - hu) f_X(q) dq du, \quad (5.5)$$

$$A_2(x) = \iint [\{g(q) - g_1(x)\}^3 + E(V^3)] L_0^3(u) f_U(x - q - hu) f_X(q) dq du, \quad (5.6)$$

$$a_r(x) = \int \{g(q) - g_1(x)\} L_0^r(u) f_U(x - q - hu) f_X(q) dq du, \quad (5.7)$$

the latter for $r = 2, 3$. Both $A_0 = \sqrt{A_0^2}$ (see (5.4)) and A_1 are nonnegative and, under Conditions A, B and C, converge to strictly positive quantities as $h \rightarrow 0$, while A_2 converges

to a finite real number. Let $\alpha \in (0, 1)$ and define $z_{1\alpha}$ and $w_{1\alpha}$ to be the solutions of $\Phi(z) = \alpha$ and $P_1(w) = \alpha$, respectively, and put $p_1(w) = A_0^{-1}a_2w^2$, $p_2(w) = \frac{1}{2}\{A_0^{-2}a_2^2(1-w^2) - A_1\}w^3$, $p_3(w) = \frac{1}{6}(A_2 + 3\sigma^2 a_3)A_0^{-3}(1-w^2)$ and $p_4(w) = \frac{1}{2}(A_0^{-1}a_2)^2(4-w^2)w^3$.

Now we are ready to establish expansions relating to the distribution of $\hat{g} - g_1$. These can be used to prove consistency of the confidence bands (see next section).

Theorem 5.1. *If Conditions A, B and C hold then*

$$\begin{aligned} \left| P_1(w) - \left\{ \Phi(w) + \delta p_1(w)\phi(w) + \delta^2 p_2(w)\phi(w) + (nh)^{-1/2}p_3(w)\phi(w) \right\} \right| \\ \leq B_1 \left\{ \delta^3 + (nh)^{-1} + \delta (nh)^{-1/2} \right\}, \end{aligned} \quad (5.8)$$

for all $n \geq n_0$, all $h \in (0, H_0]$ and all $|w| \leq B_2$, where B_1 , h and n_0 depend only on \mathcal{I} , the constants C_4, \dots, C_9 and ϵ introduced in Condition B, on the length of the interval \mathcal{I}_1 there (here and below, the constant is larger for shorter intervals), on the distributions of U and V , on the kernel K and on B_2 . If the distribution of V is either $N(0, \sigma^2)$ or gamma, as at (3.2), then dependence on the distribution of V , above, can be replaced by dependence on the constants C_1 , C_2 and C_3 in (B1).

Corollary 5.1. *If Conditions A, B and C hold then*

$$\begin{aligned} \left| w_{1\alpha} - \left\{ z_{1\alpha} - \delta p_1(z_{1\alpha}) - \delta^2 p_2(z_{1\alpha}) - (nh)^{-1/2}p_3(z_{1\alpha}) + \delta^2 p_4(z_{1\alpha}) \right\} \right| \\ \leq B_1 \left\{ \delta^3 + (nh)^{-1} + \delta (nh)^{-1/2} \right\}, \end{aligned} \quad (5.9)$$

for all $n \geq n_0$, all $h \in (0, H_0]$ and all $\alpha \in [B_2, 1 - B_2]$, where B_2 denotes any number in the interval $(0, 1)$, and B_1 , H_0 and n_0 have the dependence itemised in Theorem 5.1.

We deduce from Theorem 5.1 that

$$\begin{aligned} P\{\hat{g}(x; h) - g_1(x) \leq \delta A_0(x)w\} = \Phi(w) + \delta p_1(w)\phi(w) + \delta^2 p_2(w)\phi(w) \\ + (nh)^{-1/2}p_3(w)\phi(w) + o(\delta^2), \end{aligned} \quad (5.10)$$

where p_1 and p_3 are even, quadratic polynomials, and p_2 is an odd quintic. Since $\int K_U^2 \rightarrow \infty$ as h decreases then δ^2 is an order of magnitude larger than $(nh)^{-1}$. Therefore we have not included terms of size $(nh)^{-1}$ in (5.10). Result (5.10) fails if, on the left-hand side, we centre

$\hat{g}(x; h)$ at $g(x)$ rather than at $g_1(x)$; this goes to the heart of the impact of bias, and will be addressed in Section E.3 of the Supplementary Material.

The bound at (5.8) is applicable uniformly to distributions of X and functions g which satisfy Condition B. Therefore we can apply those bounds directly to the bootstrap setting, where F_X and g are replaced by \tilde{F}_X and \hat{g} . This allows us to consider both the conventional, unsmoothed bootstrap or its smoothed form as suggested in Section 2.3. These properties enable us to apply Theorem 5.1 directly to obtain expansions in the bootstrap case, in particular to derive

$$P\{\hat{g}^*(x; h_2) - \hat{g}_1(x) \leq \delta(h_2)\hat{A}_0(x)w|\mathcal{Z}\} = \Phi(w) + \delta(h_2)\hat{p}_1(w)\phi(w) + \delta^2(h_2)\hat{p}_2(w)\phi(w) + (nh_2)^{-1/2}\hat{p}_3(w)\phi(w) + o_p\{\delta^2(h_2)\}. \quad (5.11)$$

In fact in the bootstrap case, the analogue of the bound at (5.8) in that setting applies with probability converging to 1; and, if a constant $C > 0$ is given, then the probability converges to 1 at rate $O(n^{-C})$, provided V has sufficiently many finite moments; the number depending on C . In (5.11), hatted quantities are derived by replacing their standard forms by their bootstrap estimators; in particular, $\hat{g}_1 = E(\hat{a}^*|\mathcal{Z})/E(\hat{f}_X^*|\mathcal{Z})$, and h_2 is a bandwidth needed for constructing \hat{g}^* .

Remark 3. As a justification for using the moment-matching bootstrap when generating V_i^* s based on the first three moments, as suggested in Section 3.1, note that from (5.4) and (5.6), dependence of the polynomials p_1 , p_2 and p_3 on the distribution of V is only through the first three moments of V ; $E(V^2)$ and $E(V^3)$ appear exactly in (5.4) and (5.6), respectively. The dependence of other polynomials on the distribution of V is restricted in the same way.

5.2 Consistency of the band

The results of Section 5.1 can be used to prove pointwise consistency of the band in the ordinary smooth error case, i.e. to prove that (2.10) holds. The proof is quite standard; see Section E.3 in the Supplementary Material for details. We show there that, for the bands to have coverage error tending to zero, not all four bandwidths can be of the standard size for

estimating g . In particular, the following two settings each lead to consistent bands: (1) h_1 and $h \sim h_2$ are of conventional size for estimating g , i.e. $h_1^2 \asymp \delta(h_1)$, $h^2 \asymp \delta(h)$, $h_2^2 \asymp \delta(h_2)$, and h_0 is larger than the conventional size, so that $\delta(h_0)/h_0^2 = o(1)$; (2) h and $h_2 \sim h$ are smaller than the conventional size, that is $h^2 = o\{\delta(h)\}$ and $h_2^2 = o\{\delta(h_2)\}$, and h_0 and h_1 are such that $\delta(h_1)/h_1 = O(1)$ and $\delta(h_0)/h_0^2 = O(1)$.

Using arguments similar to those in Sections 5.5.3 and 5.5.4 of Hall (1992a), it can also be proved that, as long as, for $b = o(1)$ and $\delta(b) = o(1)$, where b denotes any of h , h_0 , h_1 , h_2 , and $h_0 \neq o(b)$, we have

$$\begin{aligned} P\{\hat{g}(x; h) - g_1(x) \leq \delta \hat{A}_0(x) \hat{w}_\alpha\} &= \Phi(z_\alpha) + \delta q_1(z_\alpha) \phi(z_\alpha) + \delta^2 q_2(z_\alpha) \phi(z_\alpha) \\ &\quad + (nh)^{-1/2} q_3(z_\alpha) \phi(z_\alpha) + o(\delta^2), \end{aligned} \quad (5.12)$$

$$\begin{aligned} P\{\hat{g}(x; h) - g_1(x) \leq -\delta \hat{A}_0(x) \hat{w}_{1-\alpha}\} &= \Phi(z_\alpha) + \delta q_1^{[1]}(z_\alpha) \phi(z_\alpha) + \delta^2 q_2^{[1]}(z_\alpha) \phi(z_\alpha) \\ &\quad + (nh)^{-1/2} q_3^{[1]}(z_\alpha) \phi(z_\alpha) + o(\delta^2). \end{aligned} \quad (5.13)$$

In (5.12) and (5.13), \hat{w}_α is defined by solving the following equation for w :

$$P\{\hat{g}^*(x; h_2) - \hat{g}_1(x) \leq \delta \hat{A}_0(x) w | \mathcal{Z}\} = \alpha, \quad (5.14)$$

and the polynomials $q_1, q_2, q_3, q_1^{[1]}, q_2^{[1]}, q_3^{[1]}$ have the same parity properties as the p_j s; they are odd or even polynomials according as the index is even or odd, respectively. See Chapter 3 in Hall (1992a) for similar results. From there, using arguments similar to, but more involved than, those in Section E.3, and the Delta method described in Section 2.7 of Hall (1992a), formulae for coverage error of the bootstrap band can be derived.

Acknowledgments

This research was supported by the Australian Research Council. We thank an associate editor and two referees for their comments which helped improve a previous version of the paper.

References

- Allison J., Santana, L. and Swanepoel, J. (2011). Two new data-dependent choices of m when applying the m -out-of- n bootstrap to hypothesis testing. *J. Statist. Comput. Simul.*, **81**, 2107–2120.

- Beran, A. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, **74**, 457–468.
- Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1095.
- Birke, M., Bissantz, N. and Holzmann, H. (2010). Confidence bands for inverse regression models. *Inverse Problems*, **26**, 115020.
- Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *J. Mult. Anal.*, **100**, 2364–2375.
- Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *J. Royal Statist. Society Ser. B.*, **69**, 483–506.
- Cao-Abad, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.*, **19**, 2226–2231.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.*, **83**, 1184–1186.
- Carroll, R. J., Ruppert, D. and Stefanski, L. A. and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models*. Chapman and Hall, UK.
- Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, **31**, 1852–1884.
- Cook, J.R. and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.*, **89**, 1314–1328.
- Cummins, D.J., Filloon, T.G. and Nychka, D. (2001). Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.*, **96**, 233–246.
- Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica*, **18**, 1025–1045
- Delaigle, A., Fan, J. and Carroll, R.J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Statist. Assoc.*, **104**, 348–359.
- Delaigle, A. and Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Stat. Data Anal.*, **45**, 249–267.
- Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Amer. Statist. Assoc.*, **103**, 280–287.
- Delaigle, A. and Hall, P. (2011). Estimation of observation-error variance in errors-in-variables regression. *Statistica Sinica*, **21**, 1023–1063.
- Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, **36**, 665–685.
- Eubank, R.L. and Speckman, P.L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.*, **88**, 1287–1301.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist.*, **19**, 1257–1272.
- Fan, J. and Truong, Y.K. (1993). Nonparametric regression with errors in variables. *Ann.*

- Statist.*, **21**, 1900–1925.
- Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.*, **16**, 433–439.
- Hall, P. (1991). On convergence rates of suprema. *Probab. Theory Related Fields*, **89**, 447–455.
- Hall, P. (1992a). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
- Hall, P. (1992b). Effects of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.*, **20**, 675–694.
- Hall, P. (1992c). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, **20**, 695–711.
- Hall, P. and Lahiri, S.N. (2008). Estimation of distribution, moments and quantiles in deconvolutions. *Ann. Statist.*, **36**, 2110–2134.
- Hall, P. and Martin, M.A. (1988). On bootstrap resampling and iteration. *Biometrika*, **75**, 661–671.
- Hall, P., and Meister, A. (2007). A Ridge-Parameter Approach to Deconvolution. *Ann. Statist.*, **35**, 1535–1558.
- Härdle, W. (1989). *Applied Nonparametric Regression*. Cambridge University Press, Boston.
- Härdle, W. and Bowman, A.W. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.*, **83**, 102–110.
- Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778–796.
- Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivar. Anal.*, **65**, 139–165.
- Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.*, **16**, 1696–1708.
- Loh, W.-Y. (1987). Calibrating confidence bands. *J. Amer. Statist. Assoc.*, **82**, 155–162.
- Mammen, E. (1992). Bootstrap, wild bootstrap, and asymptotic normality. *Probab. Theory Related Fields*, **93**, 439–455.
- Martin, M.A. (1990). On bootstrap iteration for coverage correction in confidence intervals. *J. Amer. Statist. Assoc.*, **85**, 1105–1118.
- Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canadian J. Statist.*, **32**, 439–449.
- Neumann, M.H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *Ann. Statist.*, **23**, 1937–1959.
- Neumann, M.H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparam. Statist.*, **9**, 307–333.
- Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169–184.
- Wand, M.P. and Jones, M.C. (1995). *Kernel smoothing*. Chapman and Hall, Boca Raton.
- Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.*, **14**, 1261–1350.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B*, **60**, 797–811.

A Conditions

We develop our theory under the following conditions. We choose the kernel K to satisfy:

Condition A

- (A1) K is symmetric.
- (A2) The function $(1 + x^2)|K(x)|$ is integrable.
- (A3) ϕ_K vanishes outside a compact set.
- (A4) $\phi_K(0) \neq 0$.

Write ϕ_V for the characteristic function of V . We impose the following conditions on the distributions of V and X , and on g .

Condition B

- (B1) ϕ_V satisfies Cramér’s condition, i.e. $\limsup_{|t| \rightarrow \infty} |\phi_V(t)| \leq C_1$ where $0 \leq C_1 < 1$, and also $E(V) = 0$, $E(V^2) \geq C_2 > 0$ and $E(V^4) \leq C_3 < \infty$.
- (B2) $0 < C_4 \leq \inf_{x \in \mathcal{I}} f_X(x) \leq \sup_{x \in \mathbb{R}} f_X(x) \leq C_5 < \infty$, and also f_X satisfies $|f_X(x) - f_X(x + u)| \leq C_6 |u|^\epsilon$ for all $x \in \mathcal{I}$ and all $u \in \mathbb{R}$, where $C_6 > 0$ and $0 < \epsilon \leq 1$.
- (B3) $\sup_{x \in \mathbb{R}} |g(x)| \leq C_7 < \infty$ and $|g(x) - g(x + u)| \leq C_8 |u|^\epsilon$ for all $x \in \mathcal{I}$ and all $u \in \mathbb{R}$, where $C_8 > 0$ and $0 < \epsilon \leq 1$.
- (B4) There exists a nondegenerate interval \mathcal{I}_1 , and a constant $C_9 > 0$, such that $f_X(x - q)f_U(q) \geq C_9$ for all $x \in \mathcal{I}$ and all $q \in \mathcal{I}_1$.

(The constant C_1 is referred to during proofs). Finally, recall that $L_0 = K_U / (\int K_U^2)^{1/2}$, where K_U is defined at (2.3). We make the following assumptions about the kernel, the distribution of U and the bandwidth:

Condition C

- (C1) For constants $C, H_0 > 0$, $\sup_{u \in \mathbb{R}} |L_0(u)| \leq C$ for all $h \in (0, H_0]$, and $\sup_{0 < h \leq H_0} \int_{u > u_0} L_0^2(u) du \rightarrow 0$ as $u_0 \rightarrow \infty$.
- (C2) For a function ℓ defined on the real line, and bounded away from zero on a nondegenerate interval \mathcal{I}_2 , $L_0(u) \rightarrow \ell(u)$, uniformly on all compact intervals, as $h \rightarrow 0$.

(C3) $\sup_{u \in \mathbb{R}} |f_U^{(j)}(u)| \leq C < \infty$ for $j = 0, 1$.

(C4) for some $\epsilon > 0$, $h + (nh)^{-1/2}(\int K_U^2)^{1/2} = O(n^{-\epsilon})$.

In (B1), the properties $E(V) = 0$, $E(V^2) \geq C_2 > 0$ and $E(V^4) \leq C_3 < \infty$ guarantee that the parameter k , in the gamma model at (3.2), is bounded away from zero. This ensures that the Cramér smoothness condition holds uniformly in gamma models for which $E(V) = 0$, $E(V^2) \geq C_2 > 0$ and $E(V^4) \leq C_3 < \infty$.

Next we show that (B4), (C1) and (C2) hold for commonly assumed distributions of errors, U . Indeed, if U has an ordinary smooth distribution, and in particular if $\phi_U(t) = (1 + t^2)^{-\beta/2}$ where $\beta > 1$, then for a constant $B > 0$ we have $\int K_U^2 \sim B h^{-2\beta}$. Defining $\ell_1(u) = \pi^{-1} \int_{t>0} \cos(tu) \phi_K(t) t^\beta dt$ we see that, if in addition Condition A holds, then

$$K_U(u) = \frac{1}{\pi} \int_0^\infty \cos(tu) \phi_K(t) \{1 + (t/h)^2\}^{\beta/2} dt = h^{-\beta} \ell_1(u) + o(h^{-\beta}),$$

uniformly in all real numbers u . Hence, $L_0(u) = \ell(u) + o(1)$ uniformly in u , where $\ell = B^{-1/2} \ell_1$. The function ℓ is continuous and vanishes only at isolated points, and so (C2) and the first part of (C1) holds. Note too that if K is sufficiently smooth, and $\beta > \frac{1}{2}$, then for some $\eta > 0$, $L_0(u) = O(u^{-(1/2)-\eta})$ as $u \rightarrow \infty$, from which follows the second part of (C1). Also in this setting, (B4) requires only that f_X be bounded away from zero on an open interval containing \mathcal{I} .

B Details for calibrating confidence bands

For $b = 1, \dots, B$, let \mathcal{Z}_b^* denote the bootstrap resamples and for a large number C , let $\mathcal{Z}_{b,1}^{**}, \dots, \mathcal{Z}_{b,C}^{**}$ be the double-bootstrap resamples taken from the bootstrap resample \mathcal{Z}_b^* for $b = 1, \dots, B$. These double bootstrap resamples are generated from bootstrap samples in the same way as the bootstrap samples are generated from the original sample, as in Section 2.3. The confidence bands constructed from the dataset \mathcal{Z}_b^* , $\text{CB}_{b,\hat{\beta}_\alpha}^*$, can be calculated as

$$\text{CB}_{b,\hat{\beta}_\alpha}^*(x) = [\hat{g}^*(x; h_b^*) - \hat{t}_{b,1-\hat{\beta}_\alpha/2}^*, \hat{g}^*(x; h_b^*) - \hat{t}_{b,\hat{\beta}_\alpha/2}^*], \quad b = 1, \dots, B,$$

where $\hat{g}^*(x; h_b^*)$ is the estimator of g based on the bootstrap resample \mathcal{Z}_b^* employing the bandwidth h_b^* selected by the method of Delaigle and Hall (2008). The quantity $\hat{t}_{b,\beta}^*$ is

defined as the 100β th percentile of $\hat{g}^{**}(x; h_{b,c}^{**}) - \hat{g}^*(x; h_{0,b}^*)$ with $\hat{g}^{**}(x; h_{b,c}^{**})$ constructed from the double-bootstrap resamples $\mathcal{Z}_{b,c}^{**}$ with $h_{b,c}^{**} = h_b^*$ and where $h_{0,b}^*$ is selected analogously to \hat{h}_0 , replacing, in the definition of \hat{h}_0 , \mathcal{Z} by \mathcal{Z}_b^* . We estimate $P\{g(x) \in \text{CB}_{\hat{\beta}_\alpha}^*(x) | \mathcal{Z}\}$ by

$$\frac{1}{B} \sum_{b=1}^B I\{\hat{g}(x; \hat{h}) \in \text{CB}_{b, \hat{\beta}_\alpha}^*(x)\}$$

where \hat{h} is the bandwidth calculated in Section 3.2.